Evaluation of Write-Assist Techniques for 8T SRAM in Sub-Threshold

Harsh N. Patel
Qualifying Exam
University of Virginia
hnp9uh@virginia.edu

Abstract

In an effort to operate a system at the energy optimal point, supply voltage has been scaled down to allow quadratic saving in energy according to CV_{DD}^2. However, operating the SRAM at lower supply voltages (Vmin) results in failures that necessitate an assist technique. This paper evaluates the well-known write-assist (WA) methods for an 8T SRAM cell with high threshold (V_T) devices across supply voltages. The static write margin and write delay metrics are used to evaluate different assist methods with increasing percentage of assist. The WL boosting assist increases the cell write margin (WM) up to ~7X across the supply that makes cell to write at VDD=0.3V. While Negative BL results in a ~200X reduction in write delay at VDD=0.4V with 40% assist.

Keywords

Write assist, V_{min}, sub-threshold, half-select, and high-V_T.

1. Introduction

As per the International Technology Roadmap for Semiconductors (ITRS) 2001, predicated growth of the percentage area of an embedded memory on a System-on-Chip (SoC) is ~94% by 2014 [1] and it is expected to grow in the future. Higher occupancy of the memory leads to higher energy due to the leakage that tightens constraints on memory design for low-power and energy constrained system requirements. In [2], the authors highlighted the trend of energy consumption for Body Sensor Nodes (BSNs), citing ~55% of energy consumption per operation due to memory. For energy optimal operation of an SRAM, it is required that the SRAM operates in the sub-threshold region [2], yet sub-threshold SRAM design faces its own challenges. The exponential dependence of I_{ON} on V_T in the sub-threshold region means variations will cause large yield loss due to functional failures. Also, the leakage energy begins to dominate active energy at lower supply voltages. First order standby leakage reduction can be achieved by replacing regular V_T devices with high V_T devices. In addition, the I_{ON}/I_{OFF} ratio is reduced at lower supply voltages, making 6T cell prone to read failures. Figure 1 shows the normalized I_{ON}/I_{OFF} ratio for low-V_T and high-V_T devices. The high-V_T devices have reduced I_{OFF} which results in an improvement in I_{ON}/I_{OFF} ratio in the sub-threshold region. While the same high-V_T devices faces the challenge of reduced I_{ON} and hence more failures at lower supply range.

In this paper we have considered an 8T SRAM [4] with high-V_T devices, as shown in Figure 2. The choice of an 8T cell over a 6T is due to the separate read port that mitigates

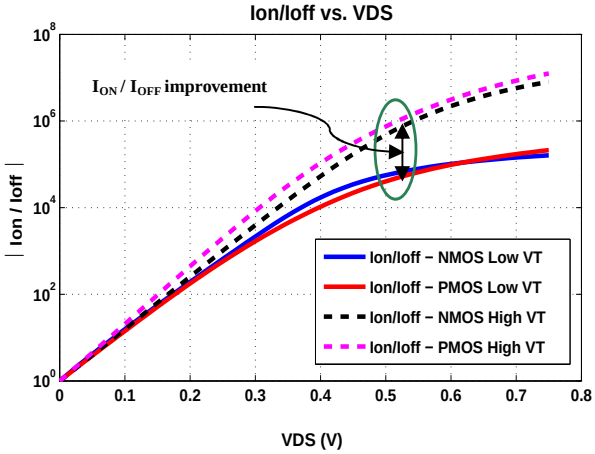

Figure 1: I_{ON}/I_{OFF} for low-V_T and high-V_T devices

read failures. For an 8T cell, read can be managed by sizing buffer separately that leads the failures in the write operation of prime concern.

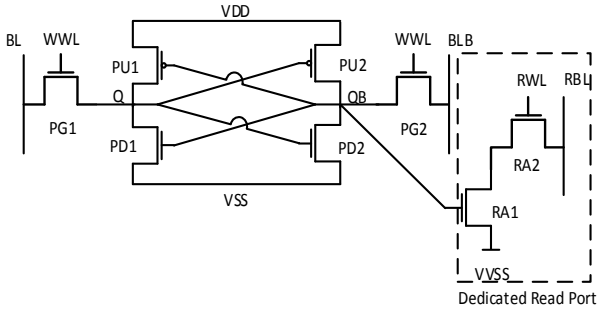

Figure 2: 8T SRAM cell

The remainder of the paper is divided as follows: Section 2 elaborates on the different assist methods considered for evaluation of the 8T bitcell and justifies the choice of the write margin and write delay as metrics for the assist techniques evaluation instead of the metrics proposed in [6]. Both metrics, WM and Write-Delay, are analyzed across different percentages of supply assist in the sub-threshold region in section 3. Section 4 concludes the paper by highlighting the optimal assist method based on the selected metric.

2. Write Assist Techniques: Theory and Prior Art

As summarized in section 1, reducing the Vmin causes write failures that necessitate the need of additional circuitry to assist the SRAM cell to operate at a lower supply range. There are various assist techniques available in the literatures, but this paper discusses four main techniques. Figure 3 shows the graphical representation of all the assist techniques.

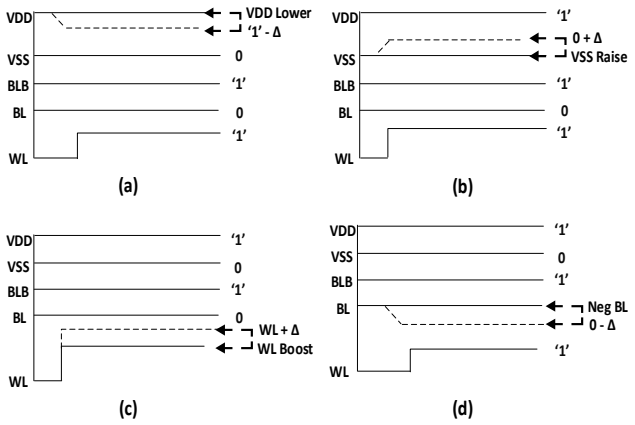

Figure 3: Write Assist Techniques: (a) VDD lower, (b) VSS raise, (c) Word-line boost, (d) Negative Bitline

With VDD lower assist, the column/core VDD is lowered to (VDD – Δ) thus reducing the $|V_{GS}|$ of the pull-up PMOS. While in VSS raise assist, the row/core VSS is raised from 0 to ΔV that weakens the PMOS by reducing its $|V_{GS}|$ due to an increase in gate voltage. Though both, VDD lower and VSS raise, weaken the PMOS, the knob that causes this change is different (source in VDD lower and gate in VSS raise) leading to their different impacts on WM. Similar to the PMOS assist techniques, word-line (WL) boost helps the pass-transistor (NMOS) by increasing the WL voltage from VDD to (VDD + Δ). This increase in V_{GS} and hence I_{ON} aids cell to flip its value easily. The WL boost scheme is implemented per row basis. In Negative Bitline (NegBL), the differential V_{GS} for an NMOS pass-transistor is raised by reducing the source voltage for the BL holding '0' by under-driving it.

In [6], the authors have considered the critical width of the WL pulse (defined as WLcrit) as an evaluation metric for the 6T cell. This dynamic metric is chosen because operating speed is the main criterion. With WLcrit metric, all WL < WLcrit is considered as failure to write for the cell. The authors also highlighted the challenge of the exponential increase in WLcrit with supply scaling as the limiting factor of the speed of operation of the SRAM. By applying different assist methods, the authors show an improvement in WLcrit, and so the speed. In [7], the authors have evaluated similar assist methods for a higher supply range, 6T cell, different technology nodes, and different metrics. Also, the authors have provided the basis for WM and write delay trade-off for choosing the optimal choice of assist at higher supply range.

To compare different assist techniques, we applied the same percentage of supply assist to all techniques. Here, we evaluate the previously mentioned assist methods with four different values of the assist: 10, 20, 30, and 40% of supply. The purpose of using different assisted values is to find the possible lower limit of Vmin using different assist techniques. For the energy constrained application space where the device operating voltage is below the threshold voltage of the device, therefore operating speed is no longer the main constraint, but the functionality. For this reason, the metric for assist evaluation is to be decided based on the SRAM usage in the application. With the WM as the static metric, the effect of each assists on timing is captured with write delay.

3. Results and Assist Comparison

To evaluate the WM and write delay metrics, we used TASE [8] for a commercial 130nm node. The results consider the impact of global variation by selecting minimum WM captured by 1000pt Monte Carlo simulation at global corner and at T=27C. The lower limit for the supply consideration has been calculated based on Data Retention Voltage (DRV), under which the cell can hold the stable states. The DRV for the cell has been calculated by sweeping the VDD till Q-QB nodes get flipped in the hold state (WL=0). With considered sizing of the 8T cell, 1000pt Monte Carlo worst-case DRV (with global variation) is 246mV. Hence VDD=0.3V is considered as the lower limit for assist evaluation.

The failures in SRAM can be divided, mainly, into three categories: write failures, read failures, and retention failures. In this paper, we limit our analysis to static write failures as they increase significantly with supply scaling. These failures for the 8T cell are measured by the write margin (WM) metric. The measurement of WM is based on the WL criteria [5], where the difference of the highest write wordline (WWL) voltage and the voltage at Q-QB nodes flipped is considered as WM.

As shown in Figure 4, the assist techniques help the cell to improve WM proportional to applied assist. In case of no assist, the cell is not able to operate below VDD=0.7V. The 10% assist brings supply voltage for the cell down to VDD=0.5V. The WL Boost assist increases the WM up to ~7X across the supply voltages and increasing applied assist range. With the increase in applied assist, the cell is able to write at VDD=0.3V using even 30% of WL boost assist technique that helps Vmin of the SRAM scale down and hence overall energy consumption. The WM trend at VDD=0.4V favors the VDD lower assist with higher percentage of supply assist, but the same faces the cell flip concern for unselected cells at deeper applied assist. The WM at 40% of supply assist with 0 values shows the cell-flip before they are being accessed. The same condition for the unselected cells limits the particular assist to go deeper.

Figure 5 summarizes each write-assist method with different percentage of supply assist applied for the write delay metric. The WM failures at VDD=0.3V except WL boost, forces the lower limit for write delay consideration at VDD=0.4V. The result shows NegBL has a higher impact on improving the write delay compared to WL Boosting and the PMOS weakening techniques across the supply range. The NegBL result extends the results claimed in [6] for the sub-threshold region. At 10% of assist, WL boosting and NegBL show similar effects on write delay. Also, the effectiveness of NegBL and WL boost assist techniques is increased as supply scales down. The difference between these two assist techniques widens as applied assist increased. The NegBL technique reduces the delay by ~200X at VDD=0.4V with 40% of supply assist.

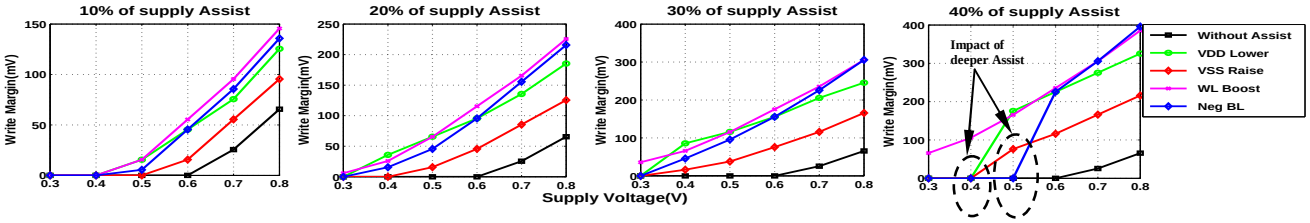

Figure 4: Assist Comparison: Write Margin metric

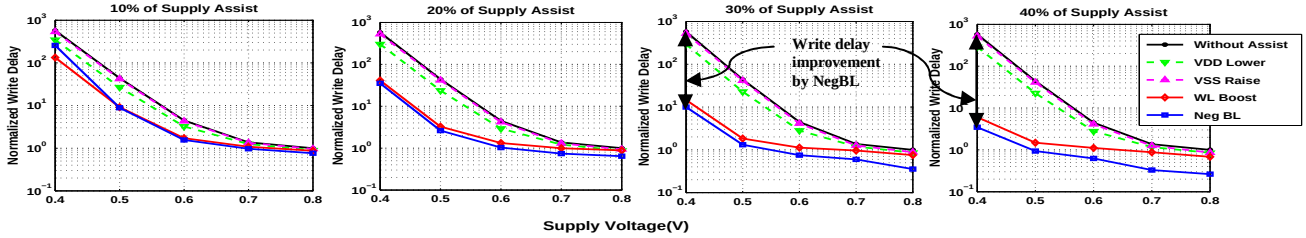

Figure 5: Assist comparison: Write Delay metric

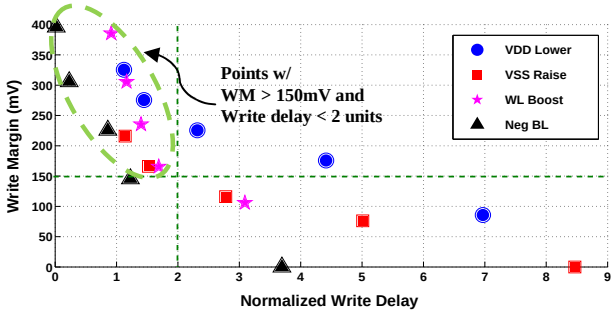

Figure 6: WM vs Write Delay at max. applied assist techniques across the supply voltages

Figure 6 consolidates the data from figure 4 and 5 with maximum applied assist techniques across the supply. These points can be combined under a constant cloud with defined WM and corresponding write delay. Such contour has been shown in the figure 6 with points having minimum WM = 150mV and maximum write delay = 2 units. With WL boost assist, required design metrics can be achieved at VDD=0.4V while other techniques requires higher supply voltages to achieve the same design metrics requirement. Similarly, based on design requirement, the selection of assist and operating supply voltage can be determined.

4. Conclusion

This paper evaluates the well-known write-assist techniques for an 8T SRAM cell with the high-V_T devices, focusing on the sub-threshold region of operation. The paper also shows the possibility to reduce the Vmin down to VDD=0.3V with sufficient WM for the write functionality using WL boosting assist. At the end, the trade-off between WM and write delay explores design decision space with possible achievable WM with corresponding write delay by each assist technique. In summary, in the sub-threshold region, the WL boosting is the optimal choice for WM metric that helps Vmin scaling requirement with Vmin down to VDD=0.3V and overall ~7X WM improvement across supply. The NegBL is optimal for the write delay with ~200X improvement at lower supply voltage.

References:

[1] J. Marinissen et al. “Challenges in Embedded Memory Design and Test”, Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE’05)

[2] Y. Zhang, et. Al., “A Batteryless 19uW MICS/ISM-Band Energy Harvesting Body Sensor Node SoC for ExG Applications,” Solid-State Circuits, IEEE Journal of, Vol. 48, no. 1, Jan. 2013.

[3] B. Calhoun, A. Wang, A. Chandrakasan, “Modeling and Sizing for Minimum Energy Operation in Subthreshold Circuit. JSSC, 2005.

[4] Verma, N.; Chandrakasan, A.P., “A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy” IEEE Journal of Solid-State Circuits, JSSC 2007.

[5] Boley, James ; Chandra, Vikas ; Aitken, Robert ; Calhoun, Benton, “Leveraging sensitivity analysis for fast, accurate estimation of SRAM dynamic write VMIN”, Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013

[6] V. Chandra, R. Aitken, C. Pietrzyk, “On the Efficacy of Write Assist Techniques in Low Voltage Nanoscale SRAMs”, DATE, 2010.

[7] R. W. Mann, J. Wang, S. Nalam, S. Khanna, G. Braceras, H. Pilo and B.H. Calhoun, “Impact of circuit assist methods on margin and performance in 6T SRAM” Solid State Electron., Nov 2010.

[8] Nalam, S., M. Bhargava, K. Ringgenberg, K. Mai, and B. H. Calhoun. “A Technology-Agnostic Simulation Environment (TASE) for Iterative Custom IC Design across Processes.” ICCD 2009.